Lecture 20

Equation-based Algorithms for CPs

Nonlinear Complementarity Problems

November 16, 2008

Outline

- Recap of Semismooth functions (Sec. 7.4)
- Application to complementarity problems
- Line search approach using FB function
- Convergence statements

Semismooth functions

- Important subclass of locally Lipschitz continuous functions
- Simple extensions of Newton's method to the nonsmooth domain using the Clarke generalized Jacobian $\partial G(x)$ do not work
- This arises because a linear model defined by $H \in \partial G(x)$ does not define a Newton approximation; Specifically, the following expression

$$\lim_{\bar{x} \neq x \to \bar{x}, H \in \partial G(x)} \frac{G(x) + H(\bar{x} - x) - G(\bar{x})}{\|x - \bar{x}\|} = 0$$

does not generally hold for any locally Lipschitzian function G.

- We illustrate this shortcoming through an example

An illustrative example

We construct a Lipschitz continuous function $f(x)$ as follows:

- We have $f(0) = 0$.

- For any integer, $n \geq 2$, let $I_n \equiv [1/n, 1/(n-1)]$ and m_n and m_{2n} be the midpoints of I_n and I_{2n}, respectively.

- Let $a_n \equiv \frac{2n}{4n-1}$ and $b_n \equiv \frac{8n-4}{4n-3}$, $a_n \leq b_n$

- Then, we define two linear functions given by $f_n^1(x) \equiv a_n(x + m_n)$ and $f_n^2(x) \equiv b_n(x - m_{2n})$.

- Furthermore, $f_n^1(\frac{1}{n-1}) = \frac{1}{n-1}$ and $f_n^2(\frac{1}{n}) = \frac{1}{n}$, $f_n^1(m_n) < f_n^2(m_{2n})$.

- Also, the point y_n defined by $f_n^1(y_n) = f_n^2(y_n)$ belongs to an open interval $(1/n, m_n)$.

- The function $f(x)$ is defined as $f(x) = \begin{cases} 0, & x = 0 \\ f_n^2(x), & x \in [1/n, y_n] \\ f_n^1(x), & x \in [y_n, 1/(n-1)] \\ f_2^1(x), & x \geq 1 \\ -f(-x), & x < 0. \end{cases}$

- Since $a_n \to \frac{1}{2}$ and $b_n \to 2$, f is locally Lipschitz on $\mathbb{R}$

- Question: **Can we construct a Newton approximation?**

- We have $\partial f(0) = [1/2, 2]$; implying that all elements of Clarke Jacobian are nonzero. Moreover, $\bar{x} = 0$ is the only solution of $f(x)$

Clarke generalized Jacobian

- To show that $f(x)+\xi(\bar{x}-x)$ with $\xi \in \partial f(x)$ is not a Newton approximation, consider the sequence $\{x^k\}$ converging to $\bar{x} = 0$, where x^k is the midpoint of $[1/k, y_k]$

- Since f is differentiable at each x^k with derivative b_k and that $x^k \leq m_k$,
$$\lim_{k\to\infty} \frac{f(x^k) + f'(x^k)(\bar{x} - x^k) - f(\bar{x})}{|x^k - \bar{x}|} = \lim_{k\to\infty} \frac{b_k x^k + b^k m_{2k} - b_k x^k}{x^k} \geq \lim_{k\to\infty} \frac{b_k m_{2k}}{m_k} = \frac{2}{3} > 0.$$
Therefore the Newton approximation definition is not met

- Consider the sequence of Newton steps defined by $x^{k+1} = x^k + \xi^k d = 0, \xi^k \in \partial f(x^k)$.

- If we start $x^0 \in (1/n, y_n)$, the function is differentiable at x^0 and is easy to check sequence of points is given by $x^1 = -m_{2n}, \quad x^2 = m_{2n}, x^3 = -m_{2n}, \ldots$

- If we start $x^0 \in (y_n, 1/(n-1))$, the function is differentiable at x^0 and is easy to check sequence of points is given by $x^1 = -m_n, \quad x^2 = m_n, x^3 = -m_n, \ldots$

- Similar behavior for $x^0 > 1$

Need for semismoothness

- Key takeaway: Taking the generalized Jacobian for a general nonsmooth function does not work

- **Semismooth functions** are locally Lipschitzian functions for which the **Clarke generalized Jacobian** provides a **legitimate Newton approximation**

Definition 1 (Defn. 7.4.2)

- Let $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ with Ω open, be a locally Lipschitz continuous function on Ω. We say that G is **semismooth** at $\bar{x} \in \Omega$ if G is directionally differentiable near $\bar{x}$ and there exists a neighborhood $\Omega' \subseteq \Omega$ of $\bar{x}$ and a function $\Delta : (0, \infty) \to [0, \infty)$ with
$$\lim_{t \to 0} \Delta(t) = 0,$$
such that for any $x \in \Omega'$ different from $\bar{x}$,
$$\frac{\|G'(x; x - \bar{x}) - G'(\bar{x}; x - \bar{x})\|}{\|x - \bar{x}\|} \le \Delta(\|x - \bar{x}\|).$$

- If the above requirement is further strengthened to
$$\frac{\|G'(x;x-\bar{x})-G'(\bar{x};x-\bar{x})\|}{\|x-\bar{x}\|^2}<\infty,$$
then we say that G is strongly semismooth at $\bar{x}$.
- If G is (strongly) semismooth at every point of Ω, then we say that G is (strongly) semismooth on Ω.

- **Insights**
  - If G is semismooth, the directional derivative $G'(\bar{x};x-\bar{x})$ provides a good approximation to $G'(x;x-\bar{x})$
  - Furthermore semismoothness of G is equivalent to (Th. 7.4.3)
    - □ $\lim_{\bar{x}\neq x\to\bar{x},H\in\partial G(x)} \frac{G'(\bar{x};x-\bar{x})-H(x-\bar{x})}{\|x-\bar{x}\|}=0;$
    - □ $\lim_{\bar{x}\neq x\to\bar{x},H\in\partial G(x)} \frac{G(x)+H(x-\bar{x})-H(x-\bar{x})}{\|x-\bar{x}\|}=0;$

- Recall that the a function is said to be **Bouligand** differentiable at a vector $x\in\Omega$, an open set, if f is Lipschitz continuous in a neighborhood of x and is directionally differentiable at x. Moreover, if f is B-differentiable, then the directional derivative along d is called the B-derivative along d.

- In fact G is semismooth at $\bar{x}$ if and only if G is B-differentiable at $\bar{x}$ and the following limit holds:

$$\lim_{\bar{x} \neq x \to \bar{x}} \frac{\|G'(x; x - \bar{x}) - G'(\bar{x}; x - \bar{x})\|}{\|x - \bar{x}\|} = 0.$$

Conversely if this limit holds, we may define $\Delta(t)$ as

$$\Delta(t) \equiv \sup_{x \in B(\bar{x},t) \setminus \{\bar{x}\}} \frac{\|G'(x; x - \bar{x}) - G'(\bar{x}; x - \bar{x})\|}{\|x - \bar{x}\|}, \quad \forall t > 0.$$

Nonlinear Complementarity Problems

- Here we consider problems of the form

$$\mathbf{0 \leq x \perp F(x) \geq 0},$$

where F is a continuously differentiable function over $\mathbb{R}^n_+$. This is one of the simplest and basic variational inequality problems

- These ideas are extendable to CPs of the form

$$\mathbf{0 \leq G(x) \perp F(x) \geq 0},$$

where F and G are continuously differentiable functions.

- Question: Why are we interested in nonsmooth formulations such as that based on a non-differentiable C-function?

- A function $\psi : \mathbb{R}^2 \to \mathbb{R}$ is called a C-function if for any pair $(a, b) \in \mathbb{R}^2$,

$$\psi(a, b) = 0 \Leftrightarrow [(a, b) \geq 0 \text{ and } ab = 0.]$$

- One reason - a smooth reformulation may fail to provide a sound basis of fast local methods?

Proposition 1 Suppose that $F : \mathbb{R}^n \to \mathbb{R}^n$ is a continuously differentiable function. Let ψ be a continuously differentiable C-function and let

$$\mathbf{F}_\psi(x) = \begin{pmatrix} \psi(x_1, F_1(x)) \\ \vdots \\ \psi(x_N, F_N(x_N)) \end{pmatrix}, \quad \forall x \in \mathbb{R}^n.$$

If x^* is a degenerate solution of NCP(F), then $J\mathbf{F}_\psi(x^*)$ is singular.

Proof: Let i be an index where $x_i^* = F_i^* = 0$. The Jacobian of $\mathbf{F}_\psi(x^*)$ is given by

$$\frac{\partial \psi(0,0)}{\partial a}(e^i)^T + \frac{\partial \psi(0,0)}{\partial b}(\nabla F_i(x^*))^T.$$

Therefore, if we can show that $\nabla \psi(0,0) = 0$, singularity of $J\mathbf{F}_\psi$ follows. But

$$\frac{\partial \psi(0,0)}{\partial a} = \lim_{a \downarrow 0} \frac{\psi(a,0)}{a} = 0,$$

since $\psi(0,0) = \psi(a,0) = 0$. (as a consequence of ψ being a C-function.) ∎

- It follows that locally fast methods may be difficult to develop using smooth C-functions (in the presence of non-degeneracy)

- More emphasis given to nonsmooth methods - appear to be better from a practical standpoint.

- **Problem:** Ensure global convergence of such nonsmooth methods

- **Important considerations**:
  - Nonsmooth equation reformulations which are (strongly) semismooth and associated merit functions are smooth
  - Desirable to use a linear Newton approximations to the nonsmooth equation - therefore so that systems of linear equations are solved at each iteration
  - Sequence of iterates should converge (globally) to solutions of CPs
  - Merit functions should have bounded level sets so that at least one limit point of the iterates exists
  - Ensure that the linear Newton approximations are nonsingular

Example

- Consider the NCP(F) given by $F : \mathbb{R} \to \mathbb{R}$ where
$$F(x) = (x-3)^2 + 1,$$
with a solution $x^* = 2$.

- Note that F is strictly monotone with JF being positive definite everywhere except at $x = 3$ where it is positive semidefinite.

- Two different unconstrained minimization reformulations
$$\theta_{\rm FB}(x) = x^2 + F(x)^2 + xF(x) - (x + F(x))\sqrt{x^2 + F(x)^2} \equiv \tfrac{1}{2}(\mathbf{F}_{\rm FB}(x))^2$$
$$\theta_{\rm MS}(x) = xF(x) + \frac{1}{4}[(x - 2F(x))_+^2 - x^2 + (F(x) - 2x)_+^2 - F(x)^2].$$
  - Both merit functions are nonnegative and continuously differentiable
  - The FB function has one stationary point (at $x = 2$) while the MS function has two ($x = 2, 3$).
  - Standard unconstrained minimization techniques will work with $\theta_{\rm FB}$ but may not work with $\theta_{\rm MS}$

Algorithms based on FB function

- We consider the equation $\mathbf{F}_{\text{FB}}(x) = 0$, where
$$\mathbf{F}_{\text{FB}}(x) \equiv \begin{pmatrix} \psi_{\text{FB}}(x_1, F_1(x)) \\ \vdots \\ \psi_{\text{FB}}(x_N, F_N(x)) \end{pmatrix},$$
where $\psi_{\text{FB}}(a,b) \equiv \sqrt{a^2+b^2} - a - b, \quad (a,b) \in \mathbb{R}^2$.

- The naturally associated merit function is given by
$$\theta_{\text{FB}}(x) \equiv \tfrac{1}{2}\mathbf{F}_{\text{FB}}(x)^T \mathbf{F}_{\text{FB}}(x).$$

- The next lemma provides some bounds for $\psi_{\text{FB}}(a,b)$ in terms of the min function and allows for showing the boundedness of the level sets

Lemma 1 (Lemma 9.1.3) *For any two scalars a and b, it holds that*
$$\frac{2}{2+\sqrt{2}}|\min(a,b)| \le |\psi_{FB}(a,b)| \le |\frac{2}{2+\sqrt{2}}\min(a,b)|.$$

Proof: Omitted.

Generalized gradient of FB function

- Example: Let $g(x) = \|x\|_2$ If $x \neq 0$, $\nabla\|x\|_2 = \frac{x}{\|x\|_2}$. Therefore $\|\nabla\|x\|_2\|_2 = 1$ for all $x \neq 0$ and Jac g(0) consists of all vectors with 2-norm of exactly one. The generalized gradient is the convex hull of this set and is given by

$$\partial\|0\| = cl\mathbf{B}(0,1).$$

- $\psi_{\mathrm{FB}}(a,b) = \sqrt{a^2+b^2} - a - b, \quad \forall (a,b) \in \mathbb{R}^2$

- Jac $\psi_{\mathrm{FB}}(0,0) = bd\mathbf{B}(0,1) - \begin{pmatrix} 1 \\ 1 \end{pmatrix} = bd\left(\begin{pmatrix} -1 \\ -1 \end{pmatrix}, 1\right).$

Differentiability properties of $\mathbf{F}_{\mathrm{FB}}$

Proposition 2 Assume that $F : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^n$ is continuously differentiable on the open set Ω. The following statements hold:

1. The generalized Jacobian of $\mathbf{F}_{\mathrm{FB}}$ satisfies

$$\partial \mathbf{F}_{\mathrm{FB}} \subseteq \mathcal{D}_a(x) + \mathcal{D}_b(x) JF(x),$$

where $\mathcal{D}_a(x)$ and $\mathcal{D}_b(x)$ are the set of $n \times n$ diagonal matrices $\mathrm{diag}(a_1(x), \ldots, a_n(x))$ and $\mathrm{diag}(b_1(x), \ldots, b_n(x))$, respectively, with

$$(a_i(x), b_i(x)) \begin{cases} \equiv \frac{(x_i, F_i(x))}{\sqrt{x_i^2 + F_i(x)^2}} - (1,1), & \text{if } (x_i, F_i(x)) \neq 0 \\ \in \ \mathrm{cl}\ \mathbf{B}(0,1) - (1,1), & \text{if } (x_i, F_i(x)) = 0. \end{cases}$$

2. $\mathbf{F}_{\mathrm{FB}}$ is semismooth on Ω.

3. θ_{FB} is continuously differentiable on Ω and its gradient $\nabla \theta_{\mathrm{FB}}$ is equal to $H^T \mathbf{F}_{\mathrm{FB}}$ for every H in $\partial \mathbf{F}_{\mathrm{FB}}(x)$

4. If the Jacobian $JF(x)$ is locally Lipshitz on Ω, then $\mathbf{F}_{\mathrm{FB}}$ is strongly semismooth on Ω.

Proof:

- **Proving (1)**
  - As a composition of Lipschitz functions, $\mathbf{F}_{\mathrm{FB}}$ is locally Lipschitz.
  - Next we restate prop. 7.1.14.

    Proposition 3 *Let $G : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$ be a locally Lipschitz on Ω an open set. If $x \in \Omega$, then $\partial G(x) \subseteq (\partial G_1(x) \times \ldots \times \partial G_m(x))^T$.*

  - It follows from this prop. that $\partial \mathbf{F}_{\mathrm{FB}}(x) \subseteq (\partial(\mathbf{F}_{\mathrm{FB}})_1(x) \times \ldots \times \partial(\mathbf{F}_{\mathrm{FB}})_n(x))^T$.
  - If i is such that $(x_i, F_i(x)) \neq (0,0)$, its easy to check that $\mathbf{F}_{\mathrm{FB}}$ is differentiable at x and

$$\nabla(\mathbf{F}_{\mathrm{FB}})_i(x) = \left(\frac{x_i}{\sqrt{x_i^2 + F_i^2(x)}} - 1\right) e^i + \left(\frac{F_i(x)}{\sqrt{x_i^2 + F_i^2(x)}} - 1\right) \nabla F_i(x),$$

where e^i is the ith unit vector

- If $(x_i, F_i(x)) = (0,0)$, we use a theorem on the generalized gradient of a composite function, recalling that $\partial\|(0,0)\| = \text{cl}B(0,1)$. Therefore

$$\nabla(\mathbf{F}_{\text{FB}})_i(x) = \{(\xi - 1)e^i + (\rho - 1)\nabla F_i(x) : (\xi, \rho) \in clB(0,1)\}.$$

Therefore (1) follows.

- **Proving (2) and (4)**
  - To prove the semismoothness of $\mathbf{F}_{\text{FB}}$, we note that each component is a composition of a semismooth function (ψ) and a smooth function $x \to (x_i, F_i(x))$. This is semismooth by prop. 7.4.4. establishes (2) and (4).

- **Proving (3)** The continuous differentiability of θ_{FB} follows from Prop. 1.5.3. It suffices to get an expression for the gradient of $\theta_{\text{FB}}(x)$. This is given by the generalized Jacobian of composite functions, implying that

$$\nabla\theta_{\text{FB}}(x) = H^T\nabla\mathbf{F}_{\text{FB}}, H \in \partial\mathbf{F}_{\text{FB}}(x).$$

Note that this result implies that $H^T\mathbf{F}_{\text{FB}}$ is independent of H if $H \in \partial\mathbf{F}_{\text{FB}}$.

Convergence analysis

We begin with a uniform continuity result on $\nabla\theta_{\mathrm{FB}}(x)$. Recall that such a notion of continuity is defined as follows.

Definition 2 *A function* $G : \mathbb{R}^n \to \mathbb{R}^m$ *is said to be uniformly continuous near a sequence* $\{x_k\} \subset \mathbb{R}^n$ *if for every* $\epsilon > 0$*, there exists a* $\delta > 0$ *such that for all* k *and all* y*,* $\|x^k - y\| \le \delta \implies \|G(x^k) - G(y)\| \le \epsilon$*.*

Lemma 2 *Let* $F : \mathbb{R}^n \to \mathbb{R}^n$ *is differentiable and with derivative* $JF : \mathbb{R}^n \times \mathbb{R}^{n\times n}$ *is uniformly continuous near a sequence* $\{x^k\}$ *and* $\{JF(x^k)\}$ *is bounded, then* F *is uniformly continuous near* $\{x^k\}$*.*

Proof omitted.

Uniform continuity of $\mathbf{F}_{\mathrm{FB}}$

Lemma 3 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be uniformly continuous near a sequence $\{x^k\} \subset \mathbb{R}^n$. The vector function $\mathbf{F}_{FB}$ is uniformly continuous near $\{x^k\}$. If in addition, $\{\mathbf{F}_{FB}(x^k)\}$ is bounded then for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all k and all matrices $D_a(y) \in \mathcal{D}_a(y)$ and $D_a(x^k) \in \mathcal{D}_a(x^k)$,*

$$\|y - x^k\| \leq \delta \implies \|[D_a(y) - D_a(x^k)]\mathbf{F}_{FB}(y)\| \leq \epsilon.$$

Proof:

- Since ψ_{FB} is a globally Lipschitz function, the uniform continuity of follows immediately. Hence for every $\epsilon > 0$, there exists a $\delta > 0$ such that for all k, i we have

$$\|y - x^k\| \leq \delta \implies |\psi_{\mathrm{FB}}(y_i, F_i(y)) - \psi_{\mathrm{FB}}(x^k, F_i(x^k))| \leq \frac{\epsilon}{8}.$$

- For simplicity, we assume that $\{\mathbf{F}_{\mathrm{FB}}(x^k)_i\}$ converges to $\mathbf{F}^{\infty}$.
  - **Case 1:** Consider an arbitrary component $\mathbf{F}_i^{\infty} = 0$, then for all sufficiently large k,

$$|(\mathbf{F}_{\mathrm{FB}}(x^k))_i| \leq \frac{\epsilon}{8}.$$

Moreover, we have

$$\begin{aligned} & |\psi_{\rm FB}(y_i, F_i(y))| \\ \leq\ & |\psi_{\rm FB}(x_i^k, F_i(x_i^k))| + |-\psi_{\rm FB}(x_i^k, F_i(x_i^k)) + \psi_{\rm FB}(y_i, F_i(y))| \\ \leq\ & \frac{\epsilon}{8} + \frac{\epsilon}{8} = \frac{\epsilon}{4}. \end{aligned}$$

- We now employ a proposition for which we provide some background. We denote diagonal matrices $D_a \in \mathcal{D}_a(x)$ and $D_b \in \mathcal{D}_b(x)$, respectively with diagonal elements given by
$$a_i(x) \equiv \xi_i - 1 \text{ and } b_i(x) \equiv \rho_i - 1,$$
for some $(\xi_i, \rho_i) \in \mathbb{R}^2$ and satisfying $\xi_i^2 + \rho_i^2 \leq 1$. It may be shown (Prop 9.1.6 b) that
$$\max(|a_i(x)|, |b_i(x)|) \leq 2.$$
- Therefore
$$|[a_i(y) - a_i(x_i^k)]\psi_{\rm FB}(y_i, F_i(y))| \leq 4\frac{\epsilon}{4} = \epsilon.$$
- **Case 2:** If $|F_i^\infty| > 0$, then by adjusting δ accordingly, from the uniform continuity of $\mathbf{F}_{\rm FB}$ near x^k for some $c > 0$, for all k sufficiently large and all y satisfying $\|y - x^k\| \leq \delta$, we have $|\psi_{\rm FB}(y_i, F_i(y))| \geq c$.

- From Lemma 9.1.3, we have that

$$\sqrt{y_i^2 + F_i(y)^2} \geq \frac{c}{2+\sqrt{2}}.$$

Hence, we deduce that

$$a_i(y) - a_i(x^k) = \frac{y_i}{\sqrt{y_i^2 + (F_i(y))^2}} - \frac{x_i^k}{\sqrt{(x_i^k)^2 + (F_i(x^k))^2}}.$$

Since both denominators are bounded away from zero and F_i is uniformly continuous near $\{x^k\}$. It follows that $|a_i(y) - a_i(x^k_)|$ can be made arbitrarily small for all k, whenever y is sufficiently close to x^k.

- Finally, by recalling that

$$\|y - x^k\| \leq \delta \implies |\psi_{\rm FB}(y_i, F_i(y)) - \psi_{\rm FB}(x^k, F_i(x^k))| \leq \frac{\epsilon}{8},$$

and that $\psi_{\rm FB}(y_i, F_i(y))$ is within a certain bound of the limit point of $\psi_{\rm FB}(x_i^k, F_i(x^k))$, the boundedness follows. Therefore $|a_i(y) - a_i(x^k)||\psi_{\rm FB}(y_i, F_i(y))|$ can be made arbitrarily small. Finally since i is arbitrary, the result follows. ∎

Uniform continuity of $\nabla\theta_{\mathbf{FB}}$

Proposition 4 *Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be a continuously differentiable function and $\{x^k\}$ be an arbitrary sequence with $\{\theta_{FB}(x^k)\}$ and $JF(x_k)$ bounded. If JF is uniformly continuous near $\{x^k\}$, then $\nabla\theta_{FB}$ is uniformly continuous near $\{x^k\}$.*

Statement of FB Line search algorithm (FBLSA)

Data: $x^0 \in \mathbb{R}^n, \rho > 0, p > 1$ and $\gamma \in (0,1)$

1. Set k $= 0$.
2. If x^k is a stationary point of $\theta_{\rm FB}$; then stop.
3. Select an element H^k in $T(x^k)$ and find a solution of the system
$$\mathbf{F}_{\rm FB}(x_k) + H^k d = 0.$$
If this system is not solvable or if
$$\nabla\theta_{\rm FB}(x^k)^T d^k \le -\rho \|d^k\|^p,$$
is not satisfied, set $d^k = -\nabla\theta_{\rm FB}(x^k)$.
4. Find the smallest nonnegative integer i_k such that with $i = i_k$
$$\theta_{\rm FB}(x^k + 2^{-i}d^k) \le \theta_{\rm FB}(x^k) + \gamma 2^{-i}\nabla\theta_{\rm FB}(x^k)^T d^k;$$
set $\tau_k = 2^{-i_k}$.
5. Set $x^{k+1} := x^k + \tau_k d^k$ and $k := k+1$; go to step 2.

Global Convergence Theory

Theorem 1 Let $F : \mathbb{R}^n \to \mathbb{R}^n$ be continuously differentiable. Let $\{x^k\}$ be an arbitrary sequence produced by (FBLSA) with T being a linear Newton approximation scheme for $\mathbf{F}_{\mathrm{FB}}$. Then the following two statements hold:

(a) Every limit point x^* of x^k satisfies $\nabla\theta_{\mathrm{FB}}(x^*) = 0$.

(b) If JF is uniformly continuous over a subsequence $\{x^k : k \in \kappa\}$, $\{JF(x^k) : k \in \kappa\}$ is bounded and $H^k \in \mathcal{D}_a(x^k) + \mathcal{D}_b(x^k)JF(x^k)$ for each $k \in \kappa$, then

$$\lim_{k(\in\kappa)\to\infty} \nabla\theta_{\mathrm{FB}}(x^k) = 0.$$

Proof: The following inequality holds for all k:

$$\|d^k\| \le \max\{\|\nabla\theta_{\mathrm{FB}}(x^k)\|, (\rho^{-1}\|\nabla\theta_{\mathrm{FB}}(x^k)\|)^{1/(p-1)}\|\}$$

and implies that if $\{\nabla\theta_{\mathrm{FB}}(x^k) : k \in \kappa\}$ is bounded then so is the set of directions given by $\{d^k : k \in \kappa\}$.

- **Proof of (a)**: Consider a subsequence given by $\{x^k : k \in \kappa\}$ which converges to x^*. We then invoke theorem 8.3.3 which is a convergence result requiring two conditions:

(BD) The objective sequence is bounded below (this would correspond to boundedness of $\nabla\theta_{\rm FB}(x^k)$)

(LS) For every sequence $\{t_k : k \in \kappa\}$ of positive scalars converging to zerom

$$\limsup_{k(\in\kappa)\to\infty} \frac{\theta(x^k + t_k d^k) - \theta(x^k) + t_k\sigma(x^k, d^k)}{t_k} \leq 0.$$

By these two statements, we may conclude that

$$\lim_{k(\in\kappa)\to\infty} \nabla\theta_{\rm FB}(x^k)^T d^k = 0.$$

Arguing as in sec. 8.3, the stationarity of x^* follows.

- **Proof of (b)**: First, we first show the boundedness of the sequence $\{\nabla\theta_{\mathrm{FB}}(x^k) : k \in \kappa\}$. For every k, there exist matrices $\widehat{D}_a(x^k)$ and $\widehat{D}_b(x^k)$ belonging to $\mathcal{D}_a(x^k)$ and $\mathcal{D}_b(x^k)$, respectively where

$$\nabla\theta_{\mathrm{FB}}(x^k) = (\widehat{D}_a(x^k) + JF(x^k)\widehat{D}_b(x^k))\mathbf{F}_{\mathrm{FB}}(x^k).$$

- By prop 9.1.6, the boundedness of the $\{\widehat{D}_a(x^k)\}$ and $\{\widehat{D}_b(x^k)\}$ follows while $\{JF(x^k) : k \in \kappa\}$ by assumption. Finally, $\mathbf{F}_{\mathrm{FB}}(x^k)$ is bounded because $\{\theta_{\mathrm{FB}}(x^k)\}$ is bounded. This establishes boundedness of $\{\nabla\theta_{\mathrm{FB}}(x^k) : k \in \kappa\}$.

- Next we verify condition (LS) from Th. 8.3.3. By boundedness of $\{\nabla\theta_{\rm FB}(x^k) : k \in \kappa\}$, so is $d^k : k \in \kappa\}$. Let $\{t_k : k \in \kappa\}$ be any sequence of scalars converging to zero and $\theta_{\rm FB}(x^k + t_k d^k) - \theta_{\rm FB}(x^k) - t_k \nabla\theta_{\rm FB}(x^k)^T d^k = t_k(\nabla\theta_{\rm FB}(x^k + t_k' d^k) - \nabla\theta_{\rm FB}(x^k))^T d^k$, for some $t_k' \in (0, t_k)$ (mean-value theoreom)

- By Prop. 9.1.9, $\nabla\theta_{\rm FB}$ is uniformly continuous near $\{x^k : k \in \kappa\}$. This coupled with boundedness of d^k for $k \in \kappa$ implies that
$$\lim_{k(\in\kappa)\to\infty} (\nabla\theta_{\rm FB}(x^k + t_k' d^k) - \nabla\theta_{\rm FB}(x^k))^T d^k = 0$$
which in turn implies that
$$\lim_{k(\in\kappa)\to\infty} \frac{\theta_{\rm FB}(x^k + t_k d^k) - \theta_{\rm FB}(x^k) - t_k \nabla\theta_{\rm FB}(x^k)^T d^k}{t_k} = 0.$$
Therefore (LS) holds.

- By invoking Theorem 8.3.3, it follows that

$$\lim_{k(\in\kappa)\to\infty} \nabla\theta_{\rm FB}(x^k)^T d^k = 0.$$

- Finally, we need to prove the statement

$$\lim_{k(\in\kappa)\to\infty} \nabla\theta_{\rm FB}(x^k) = 0.$$

  - Case 1: If $d^k = -\nabla\theta_{\rm FB}(x^k)$ infinitely many $k \in \kappa$, then the result follows.
  - Case 2: For all but finitely many $k \in \kappa$, d^k satisfies

$$\mathbf{F}_{\rm FB}(x_k) + H^k d = 0.$$
$$\nabla\theta_{\rm FB}(x^k)^T d^k \leq -\rho\|d^k\|^p,$$

    But from

$$\lim_{k(\in\kappa)\to\infty} \nabla\theta_{\rm FB}(x^k)^T d^k = 0,$$

    it follows that $d^k, k \in \kappa$ converges to zero.

  - By assumption $\{JF(x^k) : k \in \kappa\}$ is bounded while the choice of H^k implies that $\{H^k : k \in \kappa\}$ is bounded. Consequently

$$\lim_{k(\in\kappa)\to\infty} \mathbf{F}_{\rm FB}(x^k) = 0. \qquad (**)$$

- Finally, since

$$\cup_{k\in\kappa}(\mathcal{D}_a(x^k) + \mathcal{D}_b(x^k))JF(x^k)),$$

is bounded from boundedness of $\{JF(x^k) : k \in \kappa\}$ and Prop. 9.1.6. Therefore

$$\cup_{k\in\kappa}\partial\mathbf{F}_{\rm FB}(x^k)$$

is bounded.

- By prop 9.1.4(c), $\nabla\theta_{\rm FB} = H^T\mathbf{F}_{\rm FB}(x)$, where $H \in \partial\mathbf{F}_{\rm FB}(x)$. Moreover from (**), we deduce that $\{\nabla\theta_{\rm FB}(x^k) : k \in \kappa\}$ converges to zero.